

ЛЕНИВЦЕВА Ю.Д.,

Национальный Центр Когнитивных Разработок, Университет ИТМО, г. Санкт-Петербург, Россия,
e-mail: lenivezzki@gmail.com

КОПАНИЦА Г.Д.,

к.т.н., Национальный Центр Когнитивных Разработок, Университет ИТМО, г. Санкт-Петербург, Россия,
e-mail: georgy.kopanitsa@gmail.com

МЕТОД СОПОСТАВЛЕНИЯ ФОРМАТОВ ОБМЕНА МЕДИЦИНСКИМИ ДАННЫМИ И ТЕРМИНОЛОГИЙ

DOI: 10.25881/ITP.2021.79.85.007

Аннотация.

Для обеспечения непрерывной и качественной медицинской помощи важно осуществлять обмен и интеграцию медицинских данных различных провайдеров. Оценка потенциальных потерь данных в результате миграции из одной медицинской базы в другую имеет важное значение в процессе принятия решений. В данной работе описан метод предварительной оценки перекрытия пользовательских форматов с международными стандартами, а также локальных терминологических справочников с международными терминологическими системами.

Ключевые слова: медицинские записи, формат обмена данными, терминология, перекрытие.

Для цитирования: Ленивцева Ю.Д., Копаница Г.Д. Метод сопоставления форматов обмена медицинскими данными и терминологий. *Врач и информационные технологии.* 2021; 1: 75–83. doi: 10.25881/ITP.2021.79.85.007.

LENIVTCEVA I.D.,

National Center for Cognitive Technologies, ITMO University, Saint-Petersburg, Russia,
e-mail: lenivezzki@gmail.com

KOPANITSA G.D.,

PhD, National Center for Cognitive Technologies, ITMO University, Saint-Petersburg, Russia,
e-mail: georgy.kopanitsa@gmail.com

METHOD FOR MATCHING MEDICAL DATA EXCHANGE FORMATS AND TERMINOLOGIES

DOI: 10.25881/ITP.2021.79.85.007

Abstract.

Medical data exchange and integration within different care providers is highly important to ensure continuous and high-quality care service. Evaluating potential data losses during migration from one medical database to another is significant for decision-making. The presented article describes a method for preliminary overlap estimation when mapping proprietary formats with international standards, as well as local terminologies with international terminological systems.

Keywords: *medical records, data exchange format, terminology, overlap.*

How to cite: *Lenivtceva ID, Kopanitsa GD. Method for matching medical data exchange formats and terminologies. Medical doctor and information technology. 2021; 1: 75-83. (In Russ.). doi: 10.25881/ITP.2021.79.85.007.*

ВВЕДЕНИЕ

В течение жизни пациент посещает большое количество медицинских учреждений, которые собирают и хранят информацию о здоровье пациента. Данные о пациенте хранятся в различных форматах, что осложняет их интеграцию и обмен. Для качественного обмена данными необходимо обеспечить интероперабельность медицинских данных.

Понятие интероперабельности медицинских данных имеет две важные составляющие: возможность передачи данных на техническом уровне (техническая интероперабельность) и одинаковое понимание и интерпретация медицинских данных (семантическая интероперабельность). Для обеспечения интероперабельности на техническом уровне необходимо опираться на стандарты обмена медицинскими данными. Среди стандартов обмена данными в России используются следующие стандарты: openEHR [1], ISO [2], Health Level Seven v2 and v3 [3], а также Fast Healthcare Interoperability Resources (FHIR) [4]. Для обеспечения семантической интероперабельности используются международные терминологические системы такие как SNOMED CT [5], LOINC [6], ICD-10 [7] (МКБ-10).

Несмотря на широкую доступность международных стандартов, локальные терминологические системы и форматы хранения данных по-прежнему играют важную роль в медицинских системах. Обмен данными с использованием локальных форматов и терминологий представляет сложность в виду потерь данных при передаче и интеграции. Под потерями данных при обмене и интеграции мы понимаем ту часть данных, которую не удалось передать или интегрировать в силу специфики принимающего формата структурирования клинических моделей. На рисунке 1 представлен пример сопоставления данных. Параметр «Язык общения», содержащийся в МИС №1 является примером потерь данных при передаче, т.к. принимающий формат МИС №2 не предусматривает хранение этого типа данных.

Таким образом, возникает потребность сопоставления структуры локальных форматов хранения данных с международными стандартами обмена данными; и локальных терминологических систем с международными терминологиями.

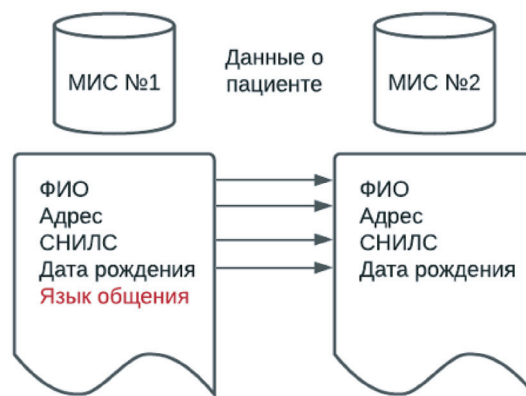


Рис. 1 – Схема сопоставления данных.

В России на сегодняшний день активно развивается стандарт обмена данными HL7 FHIR. В частности, в Санкт-Петербурге, этот стандарт успешно используется такими коммерческими медицинскими организациями как Netrika, лабораторная служба Helix, Health Samurai, Medlinx и другими компаниями, составляющими сообщество HL7 Russia.

Сопоставление форматов медицинских данных и терминологий — нетривиальная задача. Опыт сопоставления медицинских данных представлен в работах Andersen [8], Leroux [4], Doods [9], Jiang [10]. В данной работе мы применяем этот опыт к российскому случаю, чтобы оценить возможность сопоставления на примере медицинских данных из Медицинской Информационной Системы (МИС) Медиалог, и терминологических справочников из реестра НСИ Министерства здравоохранения РФ.

МЕТОДЫ

Сопоставление форматов данных

«Медиалог» — это широко распространенная коммерческая МИС, занимающая значительную долю рынка в России. По данным CNews Analytics, «Медиалог» охватывает более 20 регионов России и предоставляет услуги в рамках Единой государственной информационной системы здравоохранения (ЕГИСЗ) [11]. В основе базы «Медиалог» лежит реляционная база данных, которая не использует ни один из

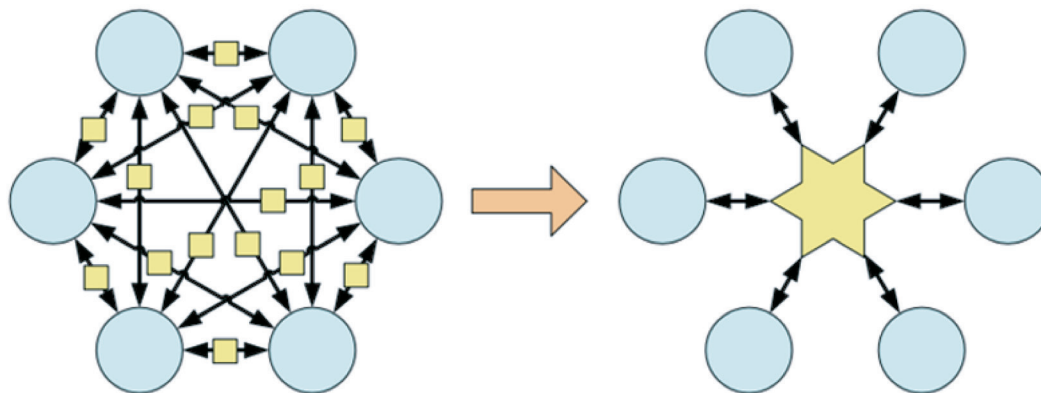


Рис. 2.

приведенных международных стандартов для структурирования медицинских данных. Этот локальный подход имеет ограничения в процедурах обмена и интерпретации информации о пациенте в результате обмена медицинскими данными. В ситуации обмена данными с другими МИС такой подход потребует разработки дополнительного интерфейса взаимодействия с каждой нестандартизированной МИС. Число таких интерфейсов растет экспоненциально с ростом числа взаимодействующих МИС (возникает ситуация комбинаторного взрыва). В случае, когда в системе присутствует стандарт, число необходимых интерфейсов равно числу МИС. Рисунок 2 наглядно демонстрирует необходимое число интерфейсов в описанных случаях. Рисунок заимствован из [12].

Также интерпретация данных осложняется использованием разных языков для представления данных, некоторые атрибуты не читаются пользователем, а некоторые записи представляют собой произвольный текст. Таким образом, система недостаточно гибкая при изменении клинических моделей и слишком сложна, чтобы обеспечить клиническую совместимость. «Медиалог» содержит множество взаимосвязанных таблиц со структурированными данными. Перекрытие между моделями данных «Медиалог» и FHIR — это та часть данных, которая может быть без потерь представлена в новом стандарте.

При сопоставлении «Медиалог» — FHIR используются восемь ресурсов FHIR, версии R4.

1. «Allergy Intolerance» описывает нежелательную реакцию организма человека на некоторое вещество.
2. «Condition» (Problem) описывает состояние здоровья пациента, диагнозы и проблемы.
3. «Family Member History» отражает важные аспекты здоровья родственника пациента.
4. «Questionnaire» — это набор вопросов, организованный для сбора ответов от пользователей/пациентов.
5. «Schedule» представляет собой расписание приемов, кабинетов, процедур и т.д.
6. «Observation» описывает простые измерения и факты о здоровье пациента.
7. «Patient» содержит административную и демографическую информацию о пациенте.
8. «Practitioner» определяет роли медицинских работников, вовлеченных в процессы медицинского учреждения.

На рисунке 3 представлена обобщенная схема сопоставления медицинских данных.

Для корректного сопоставления таблиц «Медиалог» с ресурсами FHIR мы установили ряд правил:

1. Содержимое таблицы должно быть наиболее точно выражено через ресурс. При необходимости следует использовать расширения FHIR.
2. При сопоставлении следует учитывать типы данных, количество элементов и структуры данных. Например, «datetime» из таблицы должно отображаться через «dateTime»



Рис. 3 – Схема сопоставления данных.

или другой тип данных FHIR, используемый для выражения времени или периода. «CodeableConcept» может содержать коллекцию элементов «Coding», но в ресурсе «Condition» может быть только один код.

3. Данные в формате списков из таблиц «Медиалог» должны быть сопоставлены со списками, определенными FHIR, если возможно. В противном случае следует использовать терминологические ресурсы: «NamingSystem», «CodeSystem» и «ValueSet» для создания пользовательских терминологических списков.
4. Сопоставления должны включать коды из терминологических стандартов, таких как SNOMED CT или МКБ-10.

СОПОСТАВЛЕНИЕ ТЕРМИНОЛОГИЙ

В таблице 1 приведен пример сопоставления терминов справочников НСИ с терминами SNOMED CT (Международное издание, 2019).

Сопоставление проводилось в соответствии с тремя категориями:

1. Прямое соответствие. значение исходного термина полностью покрывается целевым термином; термины взаимозаменяемы без потери информации.
2. Частичное соответствие: значения исходного и стандартного терминов не совпадают. Целевой термин можно точно описать с помощью дополнительных терминов или наоборот. Другой вариант частичного соответствия — семантическое сходство терминов.

Таблица 1 – Пример сопоставления терминологий

Термин НСИ	SNOMED код	Термин SNOMED	Категория сопоставления
Операции на органах дыхания	387661001	Operation on respiratory tract (procedure)	Прямое соответствие
Энуклеации	39250009	Enucleation (procedure)	Прямое соответствие
Операции при геморрагическом инсульте	274100004 30904006	Cerebral hemorrhage (disorder) Vascular surgery procedure (procedure)	Частичное соответствие
Эндоваскулярная тромбэкстракция	43810009	Removal of thrombus (procedure)	Частичное соответствие
Коррекция тахикардий	-	-	Соответствие отсутствует

3. Соответствие отсутствует: стандартный термин для исходного термина не найден.

В общей сложности было сопоставлено 272 термина из 8 справочников.

Проведенные сопоставления показывают возможность и полноту использования данных российской МИС «Медиалог» в международном стандарте FHIR, а также возможность и полноту использования справочников НСИ в системах, построенных на FHIR.

Оценка перекрытия форматов данных и терминологий производилась в соответствии с формулой:

$$O_{MF} = \frac{N_{ME}}{N_E} \cdot 100\%$$

где: O_{MF} — перекрытие в процентах, N_{ME} — число сопоставленных сущностей (атрибутов или терминов), N_E — общее число сущностей.

РЕЗУЛЬТАТЫ

Сопоставление форматов данных

Из таблицы 2 видно, что FHIR включает почти вдвое больше атрибутов, чем «Медиалог». Однако в некоторых случаях таблицы «Медиалог» содержат больше атрибутов, чем ресурсы FHIR, например, в таблице «Пациент».

Сопоставление терминологий

Все термины из словарей были сопоставлены вручную. В таблице 3 показано, сколько терминов из русских словарей было семантически

Таблица 2 – Расчет перекрытий ресурсов FHIR и таблиц «Медиалог»

FHIR ресурс	Кол-во атрибутов в ресурсе	Название таблицы в Медиалог	Кол-во атрибутов в таблице	Кол-во сопоставленных атрибутов Медиалог - FHIR	Кол-во сопоставленных атрибутов FHIR - Медиалог	Перекрытие Медиалог - FHIR, %	Перекрытие FHIR - Медиалог, %
Allergy Intolerance	76	DATA_SCRAT_TEST_ALLERGENS	16	13	12	81	16
Condition	80	DATA_DIAGNOSIS	46	34	32	74	40
Family Member History	67	DATA_PMT_RELATIVES_DISEASES	16	13	10	81	15
Questionnaire	109	DATA_SATISFACT_LEVEL_FORM	24	20	19	83	17
Schedule	23	PL_CABINETS_DAYS	13	11	7	85	30
Observation	121	DATA_ART_PHENOTYPE	25	23	17	92	14
Observation	121	DATA_SURGEON_EXAM	36	32	21	89	17
Observation	121	DATA_ENDOSCOPIC_EX_BIOPSY	24	20	18	83	15
Patient	91	PATIENT	192	153	27	80	30
Practitioner	58	MEDECINS_INFO	18	13	16	72	28
Итого	867		410	332	179	81	21

Таблица 3 – Сопоставление справочников НСИ и терминов SNOMED

Справочник НСИ	Прямое соответствие		Частичное соответствие		Соответствие отсутствует	
	Число	Процент	Число	Процент	Число	Процент
Виды медицинской помощи (n = 9)	1/9	11.11%	0/9	0%	8/9	88.89%
Классификатор исходов госпитализации (n = 5)	3/5	60%	2/5	40%	0/5	0%
Виды анестезии (n = 13)	10/13	76.93%	2/13	15.38%	1/13	7.69%
Типы телосложения (n = 3)	1/3	33.33%	0/3	0%	2/3	66.67%
Виды нозологических единиц диагноза (n = 7)	1/7	14.29%	1/7	14.29%	5/7	71.42%
Группы хирургических операций, проводимых в стационаре (n = 124)	89/124	71.77%	34/124	27.42%	1/124	0.81%
Классификатор исследований при остром нарушении мозгового кровообращения (n = 20)	16/20	80%	2/20	10%	2/20	10%
Классификатор фармакотерапевтических групп лекарственных средств (n = 91)	57/91	62.64%	14/91	15.38%	20/91	21.98%
Итого:	178/272	65.44%	55/272	20.22%	39/272	14.34%

отображено в SNOMED CT. Всего 178/272 (65,4%) терминов имели прямое соответствие, 55/272 (20,2%) терминов имели частичное соответствие, а 39/272 (14,3%) терминов не имели соответствия. Где n — общее количество терминов в словаре.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Сопоставление форматов данных

Расчетное перекрытие для сопоставления «Медиалог»-FHIR составляет почти 81% (таблица 3), что означает, что ресурсы FHIR покрывают приблизительно 81% атрибутов в таблицах «Медиалог». Расчетное перекрытие для сопоставления FHIR-«Медиалог» составляет 21%, что означает, что атрибуты таблиц «Медиалог» покрывают примерно 21% атрибутов ресурсов FHIR.

Многие элементы «Медиалог» повторяются и используются для внутренних сервисов «Медиалог». Таким образом, использование FHIR может улучшить процесс обмена документами за счет сокращения повторяющихся элементов. В случае сопоставления FHIR-«Медиалог» 79% потерь вызваны спецификой модели данных «Медиалог». Более того, FHIR включает дополнительную информацию, и каждое учреждение решает, какие атрибуты использовать в конкретной ситуации. В случае сопоставления «Медиалог»-FHIR потери вызваны использованием атрибутов, специфичных для «Медиалог». Однако основные элементы покрываются FHIR.

Исследование M. Andersen [8] сообщает, что 41% атрибутов Датской микробиологической модели могут быть представлены элементами FHIR, и около 7% из них актуальны в местном или национальном контексте и не имеют корреляции с ресурсами FHIR. В работе N. Hong [13] 35,1% и 40,7% составляют охват ресурсов «Medication Statement» и «Family Member History» соответственно с использованием полуавтоматического инструмента. Около 49% полученных терминов были сопоставлены с архетипами openEHR в работе J. Vuck [14] без использования расширений в университетской больнице Гейдельберга.

В исследуемых примерах сообщается, что перекрытие составляет менее 50%, что означает, что сопоставление данных локальных форматов со стандартами задача нетривиальная. Таким образом, предварительный результат для ручного сопоставления FHIR и «Медиалог» является многообещающим, а рассчитанное совпадение содержания 81% выше по сравнению с представленными результатами.

Во время сопоставления мы столкнулись со следующими проблемами:

Сравнение таблиц «Медиалог» с ресурсами FHIR вручную занимает много времени. Автоматический или полуавтоматический поиск соответствий может обеспечить более информативный анализ.

Набор атрибутов рассмотренных таблиц «Медиалог» не так информативен, как состав сопоставленных ресурсов FHIR. В среднем около 20%

информации, представленной в FHIR, отсутствует в «Медиалог».

Несмотря на существующую классификацию запросов и терминов, возникла проблема с расшифровкой некоторых атрибутов «Медиалог».

По разным причинам некоторые атрибуты, не зависящие от системы, в таблицах «Медиалог» не соответствовали определениям FHIR. Примеры таких атрибутов (см. Приложение А):

- ZUB или «Стоматологический статус». Этот атрибут не используется в системе, а поля таблицы пусты, поэтому отображение не требуется.
- PODVERGSQ_RADIACIONNOMU_O или «Пациент подвергся облучению во время аварии на Чернобыльской АЭС». Атрибут очень специфичен для российского случая и не входит в широко используемые 80% атрибутов в FHIR.
- TIP_DOPUSKA_NEDOPUSKA или «Тип недопуска» (абсолютный, относительный, временный, условный, не установлен). Поле отсутствует в ресурсе Condition с фиксированным списком кодов. И другие.

Оценка сопоставлений «Медиалог»-FHIR показала, что в России возможна стандартизация обмена данных между локальными МИС с использованием международного стандарта FHIR.

Сопоставление терминологий

Анализ результатов после сопоставления терминов показал, что только 65,4% терминов из выбранных словарей могут быть напрямую сопоставлены с SNOMED CT и использоваться в стандартизированных системах в будущем без потери информации. Около 85,7% (233/272) исследованных терминов совпадают с терминами SNOMED.

В последней работе S.H. Brown [15] получил 72% терминов, имеющих прямое соответствие в SNOMED CT при сопоставлении в ручную, 17% частичных совпадений и 11% потерь. Полученные в данной работе 65,4% прямого соответствия неудовлетворительны по сравнению с литературными результатами. Однако таблица 3

показывает, что есть словари, при сопоставлении которых возникло много потерь и выявлено мало соответствий. Мы полагаем, что справочники с высокими потерями не могут быть рекомендованы к использованию в стандартных системах и должны быть использованы локально. Справочники с низкими потерями готовы к использованию в стандартных системах. Примеры таких справочников: Классификатор исходов госпитализации (потери 0%), Виды анестезии (потери 7,7%), Группы хирургических операций, проводимых в стационаре (потери 1%). Также присутствуют справочники со средними показателями потерь. Ответственность за использование таких стандартизированных классификаторов лежит на медицинском учреждении.

ЗАКЛЮЧЕНИЕ

Чтобы понять, возможна ли стандартизация локальных форматов хранения и передачи медицинских данных, мы рассчитали перекрытие FHIR — «Медиалог». Перекрытие 81% — это хороший предварительный результат. Он показывает, что не менее 81% медицинской информации, содержащейся в МИС может быть автоматически сопоставлено с международными стандартами, в данном случае FHIR. Использование стандартизированных данных полезно как для внутренних служб обеспечения качества, так и для интеграции данных из различных источников.

Чтобы понять, можно ли использовать российские терминологические справочники в информационной системе на основе FHIR с использованием SNOMED CT, мы рассчитали процент соответствия терминов из реестра справочников НСИ терминам SNOMED CT. Исследование показало, что справочники с высоким уровнем потерь следует использовать локально.

По результатам данной работы можно сделать вывод, что присутствует высокий потенциал для стандартизации и обмена данными как между локальными МИС, так и на международном уровне. Что способствует формированию единого информационного пространства в сфере здравоохранения.

ЛИТЕРАТУРА/REFERENCES

1. Atalag K, et al. Model Driven Development of Clinical Information Systems using openEHR. *Studies in Health Technology and Informatics*. IOS Press. 2011; 169: 849–853.
2. Kopanitsa G. Evaluation Study for an ISO 13606 Archetype Based Medical Data Visualization Method. *J. Med. Syst.* 2015; 39(8): 82.
3. Rodrigues J.J.P.C, et al. *Electronic Medical Records and Their Standards*. e-Health Syst. Elsevier. 2016: 3–19.
4. Leroux H, Metke-Jimenez A, Lawley MJ. Towards achieving semantic interoperability of clinical study data with FHIR. *J. Biomed. Semantics*. 2017; 8(1): 41.
5. Kieft R.A.M.M, et al. Mapping the Dutch SNOMED CT subset to Omaha System, NANDA International and International Classification of Functioning, Disability and Health. *Int. J. Med. Inform.* Elsevier Ireland Ltd. 2018; 111: 77–82.
6. Peng P, et al. Mapping of HIE CT terms to LOINC®: analysis of content-dependent coverage and coverage improvement through new term creation. *J. Am. Med. Informatics Assoc.* 2019; 26(1): 19–27.
7. Baumel T, et al. Multi-Label Classification of Patient Notes a Case Study on ICD Code Assignment. *AAAI Conference on Artificial Intelligence*. 2017: 409–416.
8. Andersen V, et al. Feasibility of Representing a Danish Microbiology Model Using FHIR. *Stud. Health Technol. Inform.* 2017.
9. Doods J, Neuhaus P, Dugas M. Converting ODM metadata to FHIR questionnaire resources. *Studies in Health Technology and Informatics*. IOS Press. 2017; 228: 456–460.
10. Jiang G, Kiefer R, Prud'hommeaux E H.R.S. Building Interoperable FHIR-Based Vocabulary Mapping Services: A Case Study of OHDSI Vocabularies and Mappings. *Stud. Health Technol. Inform.* 2017; 245: 1327–1327.
11. Review: IT in HealthCare 2017, Informatization complicates the work of clinicians — CNews.
12. Benson T. Why Interoperability is Hard. *Principles of Health Interoperability HL7 and SNOMED*. Third Edit. 2012: 21–32.
13. Hong N, et al. Standardizing Heterogeneous Annotation Corpora Using HL7 FHIR for Facilitating their Reuse and Integration in Clinical NLP. *AMIA. Annu. Symp. proceedings. AMIA Symp. NLM (Medline)*. 2018; 2018: 574–583.
14. Buck J, et al. Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach. *Int. J. Med. Inform.* Elsevier. 2009; 78(8): 521–531.
15. Brown S.H, et al. Coverage of oncology drug indication concepts and compositional semantics by SNOMED-CT. // *AMIA. Annu. Symp. proceedings. AMIA Symp. American Medical Informatics Association*. 2003; 2003: 115–119.